

تعیین اندازه نمونه بیزی با استفاده از تابع سود نمایی مندرج به روش عددی

مسعود قاسمی بهجانی^۱، حسن زارعی^۲

تاریخ دریافت: ۱۳۹۵/۱۱/۱۶

تاریخ پذیرش: ۱۳۹۶/۱/۳۰

چکیده:

در این مقاله با پیشنهاد یک تابع سود و به دست آوردن برآورد بیزی، حجم نمونه مطلوب را بر اساس این تابع سود به دست می آوریم. این تابع سود به صورت مخصوص برای به دست آوردن برآورد بیز وقتی توزیع پسینی، گاما باشد طراحی شده است. با استفاده از این تابع سود، حجم نمونه مطلوب را برای توزیع های نرمال با میانگین معلوم، نمایی، پارتو و به قسمی می یابیم که هزینه نمونه گیری کمینه شود. در این فرایند برای کمینه کردن هزینه از تابع هزینه لیندلی استفاده می کنیم. در این جا به دلیل دشوار بودن محاسبات، اندازه نمونه را نمی توان به روش آنالیزی به دست آورد به همین دلیل به کمک روش های عددی، پواسون حجم نمونه مطلوب را به دست می آوریم.

واژه های کلیدی: توزیع پارتو، تابع هزینه، تابع سود، تابع زیان، ریسک پسینی.

۱ مقدمه

مقادیر نمونه بعدی مطرح کرده و حد بالایی را برای اندازه نمونه محاسبه نموده است. [۱] مطالعه نسبتاً جامعی روی هر دو دسته از روش ها ارائه کرده است. [۱۴] مسئله را بر اساس مقادیر مورد انتظار از اطلاعات نمونه بررسی کرده اند. [۱۹] با به کارگیری این روش در فاز دوم آزمایش های بالینی، اندازه نمونه را با ماکسیم سازی تابع سود خاصی به دست آورد. [۱۳] مسأله تعیین اندازه نمونه را با استفاده از روش بیزی برای یک آزمایش بالینی مطرح و آن را برای حالتی خاص حل کردند. [۱۱] یک تابع هدف شامل هزینه و سود اجرای یک آزمایش تصادفی را مطرح نموده، با بهینه سازی آن، اندازه نمونه آزمایش تصادفی را به روش بیزی به دست آوردند. همچنین [۱۲]، مروری بر روش های مختلف بیزی برای تعیین اندازه نمونه آزمایش های تصادفی ارائه کرد. [۱۶] در مورد تعیین اندازه نمونه در آزمایش های بالینی و حسابرسی مالی بحث کردند. در طول دهه اخیر، تعیین اندازه نمونه از دیدگاه بیزی مورد توجه بسیاری از پژوهشگران، صنعت و دولت قرار گرفته است. اگرچه هنوز بین دانشمندان بیزی و کلاسیک بحث و اختلاف نظر وجود دارد، [۴،۲] به صورت موفقیت آمیزی، شایستگی های

محاسبه اندازه نمونه در مسائل عملی موضوع مهمی است که مستقیماً با هزینه مطالعه و گردآوری داده ها ارتباط دارد. در هر پژوهش، دانستن تعداد مشاهدات (اندازه نمونه) به طوری که هزینه کلی طرح نیز مینیمم شود، مسئله ای است که از اهمیت زیادی برخوردار می باشد. هزینه کلی، شامل زیان حاصل شده از تصمیم گیری و هزینه تحلیل و هدایت آزمایش های انجام شده است. این هزینه به عوامل زیادی بستگی دارد. از جمله عوامل مهم، یکی اندازه نمونه نهایی و دیگری روش جمع آوری داده ها است. برای داشتن دقت بیشتر و در نتیجه افزایش سود باید آزمایش های بیشتری انجام شود، که این امر باعث افزایش هزینه آزمایش ها می شود. اولین مقاله بر اساس نظریه تصمیم برای تعیین اندازه نمونه [۶] بر پایه تابع زیان ارائه شد. [۷] این روش دوباره مطرح شد و مورد توجه بسیاری از پژوهشگران قرار گرفت. مسئله پیدا کردن اندازه نمونه بهینه برای هر دو نوع نمونه گیری ثابت و دنباله ای در [۳] بحث شده است. برای برآورد میانگین یک توزیع اختیاری، [۵] یک معیار بیزی را بر اساس تغییر مورد انتظار برآورد نقطه ای برای میانگین

^۱ دانش آموخته کارشناسی ارشد، گروه آمار دانشگاه سیستان و بلوچستان، ایران

^۲ عضو هیئت علمی گروه آمار دانشگاه سیستان و بلوچستان، ایران

درست‌نمایی و $p(\theta)$ تابع چگالی پیشینی برای θ باشد، چگالی پسینی برای θ به صورت زیر خواهد بود:

$$p(\theta|\mathbf{X}, n) = \frac{p(\mathbf{X}|\theta, n)p(\theta)}{p(\mathbf{X}|n)}, \quad (2)$$

که در آن $p(\mathbf{X}|n)$ تابع چگالی حاشیه‌ای X است. حال تابع سود $U(n, \mathbf{X}, \delta, \theta)$ را در نظر می‌گیریم که δ بر آورد θ است. در ابتدا از تابع سود نسبت به متغیرهای تصادفی θ و X امید ریاضی می‌گیریم و سپس بر اساس کمیت‌های δ و n سود را بیشینه می‌کنیم. با استفاده از رابطه (۲) می‌توان $p(\theta|\delta, \mathbf{X}, n)$ را به دست آورد، زیرا چگالی $p(\theta|\delta, \mathbf{X}, n)$ شبیه به $p(\theta|\mathbf{X}, n)$ است. از این رو δ تأثیری بر توزیع $\theta|\mathbf{X}, n$ ندارد. بنا بر این بیشینه‌سازی سود مورد انتظار بر اساس δ را به عنوان تصمیم نهایی اتخاذ می‌کنیم. امید ریاضی این مقدار بیشینه‌شده بر اساس \mathbf{X} را می‌توان با استفاده از $p(\mathbf{X}|n)$ به دست آورد. پس در نهایت امید ریاضی را بر حسب n بیشینه می‌کنیم تا حجم نمونه مطلوب به دست آید، که به صورت زیر می‌باشد:

$$\max_n \left[\sum_X \max_{\delta} \left\{ \int U(n, \mathbf{X}, \delta, \theta) p(\theta|\mathbf{X}, n) d\theta \right\} p(\mathbf{X}|n) \right]. \quad (3)$$

حال بر اساس [۱۵]، سود به \mathbf{X} وابسته نیست و داریم:

$$U(n, \mathbf{X}, \delta, \theta) = U(\delta, \theta) - c - cn, \quad (4)$$

که c هزینه نمونه‌گیری برای هر واحد از مشاهدات است و c هزینه پایه یا هر هزینه دیگر مربوط به نمونه‌گیری است. سپس با استفاده از روابط (۳) و (۴) داریم:

$$\max_n \left[\sum_X \max_{\delta} \left\{ \int U(\delta, \theta) p(\theta|\mathbf{X}, n) d\theta \right\} p(\mathbf{X}|n) - c - cn \right]. \quad (5)$$

در نتیجه با در نظر گرفتن روابط (۲) و (۵) حالت نهایی برای به دست آوردن حجم نمونه مطلوب به صورت زیر می‌باشد:

$$\max_n \left[\sum_X \max_{\delta} \left\{ \int U(\delta, \theta) p(\mathbf{X}|\theta, n) p(\theta) d\theta \right\} - c - cn \right]. \quad (6)$$

این روش تعیین اندازه نمونه، بیشینه‌سازی سود مورد انتظار نامیده می‌شود. در این روش، بیشینه‌سازی سود مورد انتظار دو بار مورد استفاده قرار می‌گیرد: ابتدا بر اساس θ و δ و سپس بر اساس n و X . در این مقاله از تابع زیان نسبت به تابع سود، $U(n, \mathbf{X}, \delta, \theta)$

هر دو نگرش کلاسیک و بیزی را بیان کردند. به هر حال، هیچ‌یک از بحث‌ها راجع به تعیین اندازه نمونه از دیدگاه بیزی برای آزمایش‌های بالینی در مرحله طرح‌ریزی میسر نشد؛ بنا بر این بسیاری از محققان از شبیه‌سازی استفاده کردند.

در این مقاله با در نظر گرفتن تابع هزینه لندلی و تابع سود به روش بیز به دنبال تعیین اندازه نمونه ایم به طوری که هزینه نمونه‌گیری کمینه شود. در ادامه این بخش، تعریف‌های مورد نیاز را آورده‌ایم. سپس در بخش دوم، تابع سود مورد نظر را معرفی کرده، در نهایت در بخش سوم حجم نمونه مطلوب را با استفاده از تابع سود معرفی شده به دست خواهیم آورد.

۱.۱ تابع هزینه

[۸] با فرض این که c هزینه زیرساختی از نمونه‌گیری یا هر هزینه دیگر در نمونه‌گیری باشد و c هزینه هر واحد نمونه‌گیری در نظر گرفته شود، تابع هزینه برای اندازه نمونه را به صورت

$$C(n) = c + cn; \quad n > 0 \quad (1)$$

معرفی کرد. [۱۷] با اضافه کردن ریسک پسینی بر آوردگر بیزی به این تابع هزینه، تابع هزینه کل را به صورت

$$TC(n) = c + cn + PR$$

به دست آورد.

حال با کمینه‌سازی هزینه نمونه‌گیری، حجم نمونه مطلوب را کمینه خواهیم کرد، که در برخی موارد ریسک پسینی به بردار تصادفی \mathbf{X} بستگی دارد. در این صورت امید ریاضی ریسک پسینی (APR) را به دست آورده، به تابع هزینه لندلی اضافه می‌کنیم. در نتیجه، امید ریاضی هزینه کل عبارت است از:

$$E(TC(n)) = c + cn + APR.$$

۲ تابع سود

[۹] دستورالعمل واضح تعیین اندازه نمونه بر اساس بیشترین سود مورد انتظار^۳ را با در نظر گرفتن تابع هزینه خطی به صورت زیر تعریف کرد:

فرض کنید X_1, X_2, \dots, X_n یک نمونه n تایی از تابع چگالی $p(X|\theta)$ باشد که پارامتر θ مجهول است. اگر $p(\mathbf{X}|\theta, n)$ تابع

^۳ Maximization of Expected Utility (MEU)

۲.۲ تابع سود نمایی مندرج

اگر δ برآورد θ باشد تابع سود را به صورت زیر تعریف می‌کنیم:

$$U(\delta, \theta) = \frac{\theta}{\delta} \exp \left[\left(1 - \frac{\theta}{\delta} \right) \right], \quad (9)$$

که $\theta \in R_+$ و سود در نقطه $\theta = \delta$ بیشینه می‌شود. به عبارت دیگر $U(\delta, \delta) = 1$. متناوباً می‌توانیم تابع سود بالا را به عنوان تابع زیان به صورت زیر در نظر بگیریم:

$$L(\delta, \theta) = 1 - r \exp(1 - r), \quad (10)$$

که $r = \frac{\theta}{\delta}$ و $r = 1$ یعنی $\delta = \theta$ ؛ به این مفهوم که هیچ زبانی نداریم یا به عبارت دیگر $L(\delta, \delta) = 0$.

تابع زیان (۱۰) حالت خاصی از تابع زیان گامای بازتابی^۴ با شرط $\gamma = 1$ است. تابع زیان گامای بازتابی توسط [۱۸] معرفی شد و به صورت زیر است:

$$L(\theta, \delta) = k \left[1 - \left(\frac{\delta}{\theta} \right)^\gamma e^{-\gamma \left(\frac{\delta}{\theta} - 1 \right)} \right], \quad k > 0, \gamma > 0.$$

که در آن k بیشینه زیان و γ پارامتر شکل می‌باشد. خواص:

(۱) این تابع زیان، کراندار است و بیشینه زیان حد اکثر مقدار یک و کمینه زیان مقدار صفر را می‌گیرد (حداقل مقدار زیان، صفر است، زیان منفی نمی‌شود).

(۲) زیان کاهش می‌یابد اگر $1 < r < \infty$.

(۳) اگر $r = 1$ باشد، زیان برابر با صفر است و خطای برآورد $r - 1$ می‌باشد.

(۴) زیان به ازای مقادیر $r > 1$ افزایش می‌یابد.

این تابع زیان نامتقارن است، زیرا $L(\delta, \theta)$ به ازای $0 < r < 1$ به سرعت کاهش می‌یابد و برای $r = 1$ برابر با صفر می‌باشد.

در نهایت به ازای $r > 1$ زیان به طور آهسته افزایش پیدا می‌کند.

از این رو صورت نمایی سود به وسیله $r = \frac{\theta}{\delta}$ مندرج (به ازای مقادیر مختلف r زبان‌های متفاوتی به دست می‌آید، $r = 1$ زیان صفر، $r > 1$ زیان افزایش و $0 < r < 1$ زیان کاهش می‌یابد)

می‌شود. به همین دلیل تابع سود پیشنهادی را تابع سود نمایی مندرج نامیده‌ایم.

استفاده می‌کنیم که زیان به صورت زیر است:

$$-U(n, \mathbf{X}, \delta, \theta) = L(\delta, \theta) + c. + cn. \quad (V)$$

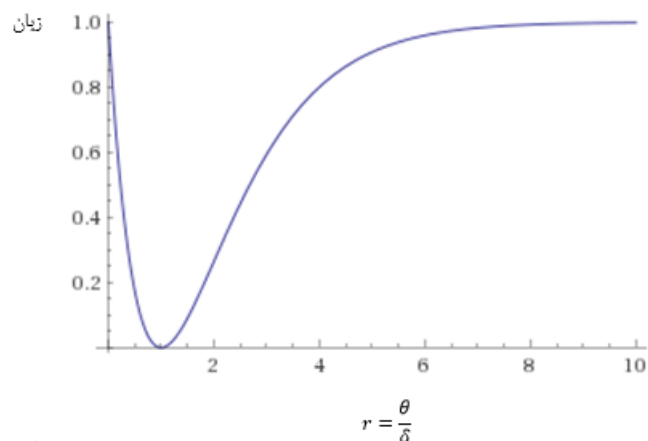
و صورت نهایی برای به دست آوردن کوچک‌ترین اندازه نمونه عبارت خواهد بود از

$$\min_n \left[\sum_X \max_\delta \left\{ \int L(\delta, \theta) p(\mathbf{X}|\theta, n) p(\theta) d\theta \right\} + c. + cn \right]. \quad (8)$$

۱.۲ دستورالعمل

در این مقاله با معرفی تابع سود و برآورد بیز، حجم نمونه مطلوب را بر اساس تابع سود معرفی شده به دست می‌آوریم. این تابع سود به صورت مخصوص برای به دست آوردن برآورد بیز وقتی توزیع پسینی، گاما باشد طراحی شده است. اگر δ برآورد θ باشد تابع سود و زیان باید تابعی از δ و θ باشند، که ترکیب δ و θ به صورت تفاضل یا نسبتی از آنها یا تفاضل و نسبت می‌باشد. اما بین توابع سود و زیان، رابطه عکس وجود دارد. اگر تابع زیان را با $L(\delta, \theta)$ و تابع سود را با $U(\delta, \theta)$ تعریف کنیم، و با توجه به این که زیان عبارت است از بیشترین سود قابل کسب منهای سودی که در عمل به دست می‌آید، رابطه بین تابع سود و زیان به صورت $L(\delta, \theta) = \max_\theta U(\delta, \theta) - U(\delta, \theta)$ می‌باشد؛ به این معنی که برای برآورد بیز منحصر به فرد، کمینه‌سازی زیان مورد انتظار معادل است با بیشینه‌سازی سود مورد انتظار. در عمل می‌خواهیم برای تصمیم‌ها به طور دقیق توابع سود و زیان را در نظر بگیریم. در بیشتر موارد از میانگین پسینی استفاده می‌شود. بنا بر این اگر میانگین پسینی برآوردی برای انتخاب باشد، پس تابع سود یا زیان به صورت درجه دوم است. حال برای این که نشان دهیم می‌توان هر تابع سود یا زبانی که نامتقارن است و میانگین پسینی را به عنوان برآورد بیز در نظر می‌گیرد داشته باشیم، یک تابع سود برای پارامتر عضو اعداد حقیقی مثبت تعریف می‌کنیم و در ادامه سود مورد انتظار را برای این تابع سود به دست می‌آوریم. بنا بر این یک صورت هم‌ارز تابع زیان ارائه می‌کنیم و در نهایت حجم نمونه مطلوب را با کمینه کردن زیان مورد انتظار به دست می‌آوریم. بیشتر توابع ریسک پسینی، صورت‌هایی پیچیده دارند و نمی‌توان به سادگی و با مشتق‌گیری حجم نمونه مطلوب را به دست آورد که در این حالات، حجم نمونه مطلوب را با استفاده از روش‌های عددی به دست می‌آوریم.

^۴ Rlected gamma loss function



شکل ۱. تابع زیان نمایی مندرج به ازای $r = \frac{\theta}{\delta}$

در قضیه زیر، برآورد بیز را برای تابع سود معرفی شده به دست می آوریم.

$$\delta = \frac{\alpha}{\beta}$$

و در نهایت داریم:

که به وضوح میانگین پسینی است.

قضیه ۲.۲. فرض کنید X_1, X_2, \dots, X_n یک نمونه تصادفی به اندازه n از هر توزیعی با تابع چگالی $p(X|\theta)$ باشد که $\theta \in R_+$ همچنین فرض کنید $p(\theta)$ هر تابع چگالی پیشینی از توزیع θ باشد به طوری که توزیع پسینی با تابع چگالی $\varphi(\theta|\mathbf{X})$ توزیع گاما با پارامترهای α و β باشد. پس پیشینه سود مورد انتظار بر اساس تابع سود (۹)، $e^{-1} \times \left(1 + \frac{1}{\alpha}\right)^{-(\alpha+1)}$ است.

اثبات. فرض کنید X_1, X_2, \dots, X_n یک نمونه تصادفی به اندازه n از تابع چگالی به صورت $p(X|\theta)$ است. اگر $p(\theta)$ تابع چگالی پیشینی برای θ باشد، به طوری که توزیع پسینی با چگالی $\varphi(\theta|\mathbf{X})$ توزیع گاما با پارامترهای α و β باشد، از قضیه ۱.۲ برآورد بیز بر اساس تابع سود (۹) برابر است با $\delta = \frac{\alpha}{\beta}$ ، که میانگین پسینی است. بنا بر این سود مورد انتظار به صورت زیر به دست می آید:

$$E[U(\delta, \theta)] = \int \frac{\theta}{\delta} \exp\left(1 - \frac{\theta}{\delta}\right) p(\theta|\mathbf{X}) d\theta,$$

که در نقطه $\delta = \frac{\alpha}{\beta}$ بیشینه می شود. پس داریم:

$$E[U(\delta, \theta)] = e \times \left(1 + \frac{1}{\alpha}\right)^{-(\alpha+1)}.$$

این عبارت به وضوح مستقل از پارامتر پیشینی β است. در نتیجه معادل صورت ریسک پسینی بر اساس تابع زیان (۱۰) به شکل زیر می باشد:

$$PR = 1 - e \times \left(1 + \frac{1}{\alpha}\right)^{-(\alpha+1)}. \quad (12)$$

قضیه ۱.۲. فرض کنید X_1, X_2, \dots, X_n یک نمونه n تایی از هر توزیعی با تابع چگالی $p(X|\theta)$ باشد که $\theta \in R_+$ همچنین فرض کنید $p(\theta)$ هر تابع چگالی پیشینی برای توزیع θ است به طوری که توزیع پسینی با تابع چگالی $\varphi(\theta|\mathbf{X})$ توزیع گاما با پارامترهای α و β باشد. پس برآورد بیز بر اساس تابع سود (۹) میانگین پسینی یعنی $\frac{\alpha}{\beta}$ است.

اثبات. فرض کنید X_1, X_2, \dots, X_n یک نمونه تصادفی به اندازه n از تابع چگالی به صورت $p(X|\theta)$ باشد. اگر $p(\theta)$ تابع چگالی پیشینی برای θ باشد، توزیع پسینی، گاما با پارامترهای α و β است. حال از تابع سود (۹) نسبت به $p(\theta|\mathbf{X})$ امید ریاضی می گیریم:

$$E[U(\delta, \theta)] = \int \frac{\theta}{\delta} \exp\left(1 - \frac{\theta}{\delta}\right) p(\theta|\mathbf{X}) d\theta \quad (11)$$

سپس از رابطه (۱۱) نسبت به δ مشتق می گیریم:

$$\begin{aligned} \frac{\partial E[U(\delta, \theta)]}{\partial \delta} &= \frac{\beta^\alpha \exp(1)}{\Gamma(\alpha)} \int \theta^{\alpha+1-1} \exp\{-\beta\theta\} \\ &\times \left[\frac{1}{\delta} \left(\frac{\theta}{\delta^2}\right) \exp\left\{-\frac{\theta}{\delta}\right\} - \frac{1}{\delta^2} \exp\left\{-\frac{\theta}{\delta}\right\} \right] d\theta \\ &= \frac{\alpha \beta^\alpha \exp(1)}{\delta^2 (\beta + \frac{1}{\delta})^{\alpha+1}} \left[\frac{\alpha + 1}{\delta (\beta + \frac{1}{\delta})} - 1 \right]. \end{aligned}$$

برای به دست آوردن برآورد بیز، عبارت $\frac{\partial E[U(\delta, \theta)]}{\partial \delta}$ را برابر با صفر قرار می دهیم:

$$\frac{\alpha \beta^\alpha \exp(1)}{\delta^2 (\beta + \frac{1}{\delta})^{\alpha+1}} \left[\frac{\alpha + 1}{\delta (\beta + \frac{1}{\delta})} - 1 \right]$$

می‌یابد و به‌خاطر ناچیز بودن قابل چشم‌پوشی است) داریم:
 $PR \approx 1 - \exp\left(-\frac{1}{z}\right)$ با جایگذاری مقدار z در PR داریم:

$$PR \approx 1 - \exp\left(-\frac{1}{\alpha + \frac{n}{\frac{t}{\beta}} + 1}\right)$$

حال با اضافه کردن ریسک پسینی به تابع هزینه (۱)، هزینه کل به‌صورت زیر به دست می‌آید:

$$TC(n) \approx c + cn + 1 - \exp\left(-\frac{1}{\alpha + \frac{n}{\frac{t}{\beta}} + 1}\right) \quad (13)$$

برای به دست آوردن حجم نمونه مطلوب از رابطه (۱۳) نسبت به n مشتق گرفته، برابر با صفر قرار می‌دهیم، که در نتیجه داریم:

$$c\left(\alpha + \frac{n}{\frac{t}{\beta}} + 1\right)^2 - \exp\left\{-\frac{1}{\frac{t}{\beta}\left(\alpha + \frac{n}{\frac{t}{\beta}} + 1\right)^2}\right\} = 0.$$

حل این معادله به‌صورت تحلیلی امکان‌پذیر نیست؛ به همین دلیل از روش عددی استفاده می‌کنیم. با استفاده از نرم‌افزار Maple13 و حل معادله بالا بر حسب n ، دو ریشه به دست می‌آید که فقط یکی از ریشه‌ها مثبت است و باید همان ریشه مثبت را به‌عنوان حجم نمونه مطلوب برای مقادیر c و α در نظر گرفت. در جدول ۱ حجم نمونه مطلوب را برای مقادیر مختلف c با فرض ثابت بودن α و در جدول ۲ حجم نمونه مطلوب را با فرض ثابت بودن c برای مقادیر مختلف α به دست آورده‌ایم.

که $z = \alpha + n + 1$. حال با بسط دادن عبارت لگاریتم و صرف نظر کردن از توان سوم و بالاتر z ، داریم $PR \approx 1 - \exp\left(-\frac{1}{z}\right)$. بنا بر این با جایگذاری مقدار z ، ریسک پسینی به‌صورت زیر به دست می‌آید:

$$PR \approx 1 - \exp\left(-\frac{1}{\alpha + n + 1}\right) \quad (14)$$

حال ریسک پسینی را به تابع هزینه (۱) اضافه می‌کنیم تا هزینه کل به‌صورت زیر به دست می‌آید:

$$TC(n) \approx c + cn + 1 - \exp\left(-\frac{1}{\alpha + n + 1}\right) \quad (15)$$

۳ تعیین اندازه نمونه

در این بخش حجم نمونه مطلوب را با استفاده از قضایای بیان‌شده برای توزیع‌های نرمال با فرض معلوم بودن میانگین، نمایی، پارتو و پواسون به دست می‌آوریم.

۱.۳ تعیین اندازه نمونه در برآورد دقت توزیع نرمال وقتی میانگین معلوم است

فرض کنید X_1, X_2, \dots, X_n یک نمونه تصادفی به اندازه n از توزیع نرمال با میانگین μ و دقت $\xi (= \frac{1}{\sigma^2})$ باشد. با فرض این که توزیع پیشینی، توزیع مزدوج گاما با پارامترهای α و β است، توزیع پسینی نیز گاما می‌باشد با پارامترهای $\alpha + \frac{n}{\frac{t}{\beta}}$ و $\beta + \frac{t}{\frac{t}{\beta}}$ که $T = \sum_{i=1}^n (X_i - \mu)^2$ در نتیجه طبق قضیه ۱.۲، برآورد بیز بر اساس تابع زیان (۱۰)، به‌صورت $\delta = \frac{\alpha + \frac{n}{\frac{t}{\beta}}}{\beta + \frac{t}{\frac{t}{\beta}}}$ است. در ادامه و با استفاده از قضیه ۲.۲، امید ریاضی تابع سود به‌صورت زیر به دست می‌آید:

$$E[U(\xi, \theta)] = e \times \left(1 + \frac{1}{\alpha + \frac{n}{\frac{t}{\beta}}}\right)^{-(\alpha + \frac{n}{\frac{t}{\beta}} + 1)}$$

بنا بر این بر اساس رابطه (۱۲)، تابع ریسک پسینی برابر است با:

$$PR = 1 - \exp\left\{1 + z \log\left(1 - \frac{1}{z}\right)\right\}$$

که $z = \alpha + \frac{n}{\frac{t}{\beta}} + 1$. با بسط دادن عبارت لگاریتم و صرف نظر از توان سوم و بالاتر z (زیرا با افزایش توان، $\frac{1}{z}$ کاهش

۲.۳ تعیین اندازه نمونه در برآورد پارامتر توزیع نمایی

فرض کنید X_1, X_2, \dots, X_n یک نمونه تصادفی به اندازه n از توزیع نمایی با پارامتر λ باشد. همچنین فرض کنید توزیع مزدوج پیشینی، گاما با پارامترهای α و β باشد. پس توزیع پسینی نیز گاما است با پارامترهای $\alpha + n$ و $\beta + s$ که $S = \sum_{i=1}^n X_i$ بنا بر این طبق قضیه ۱.۲، برآورد بیز بر اساس تابع زیان (۱۰) به‌صورت $\delta = \frac{\alpha + n}{\beta + s}$ است و بر اساس رابطه (۱۲) ریسک پسینی به‌صورت زیر می‌باشد:

$$PR = 1 - \exp\left\{1 + z \log\left(1 - \frac{1}{z}\right)\right\},$$

جدول ۱. حجم نمونه مطلوب در برآورد دقت توزیع نرمال با میانگین معلوم به ازای مقادیر مختلف c

c	n^*		
	$\alpha = 1$	$\alpha = 2$	$\alpha = 3$
۰/۰۰۰۵	۴۰	۳۶	۳۳
۰/۰۰۱۰	۳۶	۳۲	۲۹
۰/۰۰۱۵	۳۳	۲۹	۲۵
۰/۰۰۲۰	۲۸	۲۴	۲۱
۰/۰۰۲۵	۲۴	۲۰	۱۷
۰/۰۰۳۰	۱۹	۱۷	۱۴
۰/۰۰۳۵	۱۶	۱۴	۱۰
۰/۰۰۴۰	۱۴	۱۱	۶
۰/۰۰۴۵	۱۱	۸	۳
۰/۰۰۵۰	۷	۶	۱

جدول ۲. حجم نمونه مطلوب در برآورد دقت توزیع نرمال با میانگین معلوم به ازای مقادیر مختلف α

α	n^*		
	$c = 0/001$	$c = 0/0015$	$c = 0/002$
۱	۳۶	۳۳	۲۸
۲	۳۲	۲۹	۲۴
۳	۲۹	۲۵	۲۱
۴	۲۵	۲۱	۱۸
۵	۲۲	۱۹	۱۴
۶	۱۹	۱۵	۱۱
۷	۱۶	۱۱	۷
۸	۱۳	۹	۵
۹	۱۱	۶	۲
۱۰	۸	۴	۱

ثابت بودن α و در جدول ۴ حجم نمونه مطلوب را برای مقادیر مختلف پارامتر پیشینی α با شرط ثابت بودن c به دست آورده‌ایم.

و در نهایت برای به دست آوردن حجم نمونه مطلوب از رابطه (۱۵) نسبت به n مشتق گرفته، برابر با صفر قرار می‌دهیم، که داریم:

۳.۳ تعیین اندازه نمونه در برآورد پارامتر توزیع پارتو

$$2c(\alpha + n + 1)^2 - \exp\left\{-\frac{1}{2(\alpha + n + 1)^2}\right\} = 0 \quad (16)$$

فرض کنید X_1, X_2, \dots, X_n یک نمونه تصادفی به اندازه n از توزیع نمایی با پارامتر θ باشد، که دارای تابع چگالی به صورت زیر است:

حال با استفاده از نرم افزار Maple13 رابطه (۱۶) را بر حسب n حل می‌کنیم. در ادامه دو ریشه به دست می‌آید که فقط یکی مثبت است و آن را به عنوان حجم نمونه مطلوب برای مقادیر مختلف c و α در نظر می‌گیریم. در جدول ۳ حجم نمونه مطلوب را در برآورد پارامتر توزیع نمایی برای مقادیر مختلف c با فرض

$$p(X|\theta) = \theta \alpha^\theta x^{-(\theta+1)}; \quad x > \alpha, \theta > 0$$

و می توان آن را به شکل زیر نوشت:

$$p(X|\theta) = \theta \exp \left\{ -\theta \log \left(\frac{x}{\alpha} \right) \right\} \frac{1}{x}$$

(۱۲)، ریسک پسینی به صورت زیر به دست می آید:

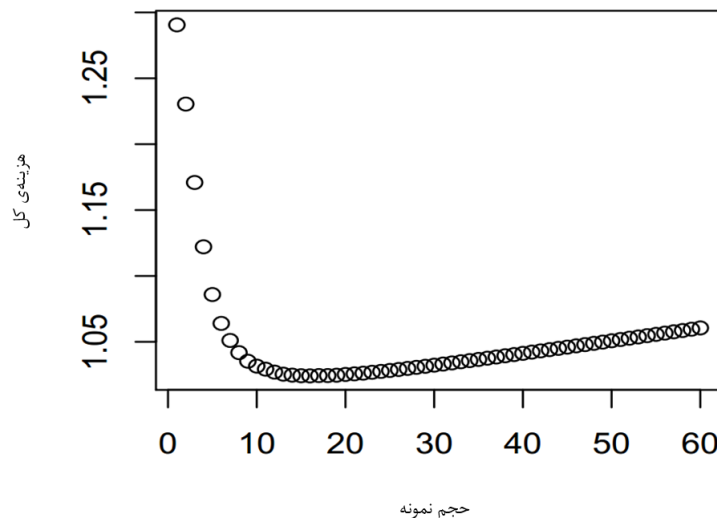
$$PR = 1 - \exp(-1) \times \left(1 + \frac{1}{\alpha+n} \right)^{-(\alpha+n+1)}$$

همچنین فرض کنید توزیع مزدوج پیشینی برای θ ، توزیع گاما با پارامترهای α و β باشد، که در نتیجه توزیع پسینی نیز گاما است با پارامترهای $\alpha+n$ و $\beta + \sum_{i=1}^n \log x_i - n \log \alpha$ در نتیجه بر اساس قضیه ۱.۲، برآورد بیز بر اساس تابع زیان (۱۰) به صورت

که دقیقاً شبیه به ریسک پسینی در رابطه (۱۴) است. پس ریسک پسینی بر اساس تابع زیان (۱۰) مستقل از پارامتر پیشینی β می باشد.

جدول ۳. حجم نمونه مطلوب در برآورد پارامتر توزیع نمایی به ازای مقادیر مختلف c

c	n^*		
	$\alpha = 1$	$\alpha = 3$	$\alpha = 5$
۰/۰۰۰۱۰	۷۴	۶۹	۶۵
۰/۰۰۰۱۵	۶۸	۶۱	۵۲
۰/۰۰۰۲۰	۶۲	۵۵	۴۴
۰/۰۰۰۲۵	۵۷	۴۹	۳۹
۰/۰۰۰۳۰	۵۱	۴۴	۳۵
۰/۰۰۰۳۵	۴۶	۴۰	۳۲
۰/۰۰۰۴۰	۴۰	۳۴	۲۹
۰/۰۰۰۴۵	۳۶	۳۱	۲۶
۰/۰۰۰۵۰	۳۲	۲۸	۲۲
۰/۰۰۰۵۵	۲۹	۲۵	۱۹
۰/۰۰۰۶۰	۲۷	۲۳	۱۵
۰/۰۰۰۶۵	۲۴	۱۹	۱۱
۰/۰۰۰۷۰	۲۲	۱۶	۸



شکل ۲. تعیین اندازه نمونه در برآورد پارامتر توزیع پواسون به ازای $\alpha = 1, \beta = 4, c = 1$ و $c = 0/001$.

جدول ۴. حجم نمونه مطلوب در برآورد پارامتر توزیع نمایی به ازای مقادیر مختلف α

α	n^*		
	$c = 0.0001$	$c = 0.00015$	$c = 0.0002$
۵	۶۵	۴۴	۲۲
۱۰	۵۹	۳۹	۲۱
۱۵	۵۴	۳۶	۱۸
۲۰	۵۰	۳۱	۱۶
۲۵	۴۶	۲۷	۱۳
۳۰	۴۳	۲۴	۱۱
۳۵	۳۹	۲۲	۷
۴۰	۳۴	۱۹	۵
۴۵	۳۱	۱۷	۲
۵۰	۲۸	۱۴	۱

جدول ۵. حجم نمونه مطلوب در برآورد پارامتر توزیع پواسون برای مقادیر مختلف پارامترهای پیشینی α و β به ازای $c = 1$ و

$c = 0.001$

α	β	n^*	β	α	n^*
۱	۱	۱۰	۱	۱	۱۰
	۲	۱۲		۲	۶
	۳	۱۴		۳	۴
	۴	۱۵		۴	۲
۲	۱	۶	۲	۱	۱۲
	۲	۸		۲	۸
	۳	۱۰		۳	۶
	۴	۱۲		۴	۴
۳	۱	۴	۳	۱	۱۴
	۲	۶		۲	۱۰
	۳	۸		۳	۸
	۴	۱۰		۴	۶

۴.۳ تعیین اندازه نمونه در برآورد پارامتر توزیع پواسون

فرض کنید X_1, X_2, \dots, X_n یک نمونه تصادفی به اندازه n از توزیع نمایی با پارامتر θ باشد. همچنین فرض کنید توزیع پیشینی برای θ ، گاما با پارامترهای α و β باشد. پس توزیع پسینی نیز گاما است با پارامترهای $\beta + t$ و $\alpha + n$ ، که $T = \sum_{i=1}^n X_i$ بر اساس قضیه ۱.۲، برآوردگر بیز بر اساس تابع زیان (۱۰)

بنا بر این در این حالت دقیقاً تابع هزینه (۱۵) را داریم که در نتیجه برای به دست آوردن حجم نمونه مطلوب باید رابطه (۱۶) را حل کنیم، که همان نتایج عددی حاصل از توزیع نمایی به دست می آید.

نتیجه گیری

در این مقاله حجم نمونه مطلوب را برای توزیع های مختلف بر اساس تابع سود محاسبه کردیم. در این جا یک تابع سود نامتقارن جدید پیشنهاد کردیم. مزیت استفاده از این تابع سود، این است که کراندار است و برای توزیع پسینی گاما، میانگین پسینی همان برآورد بیز می باشد. همچنین امید ریاضی این تابع سود با توزیع مزدوج پیشینی برای این صورت نامتقارن از سود را به دست آوردیم و در ادامه اندازه نمونه مطلوب را با استفاده از این تابع سود برای توزیع های نرمال با میانگین معلوم، نمایی، پارتو و پواسون محاسبه کردیم. حجم نمونه مطلوب را برای توزیع های نرمال، نمایی و پارتو در جدول هایی ارائه دادیم که برای این توزیع ها با افزایش پارامترهای پیشینی، حجم نمونه مطلوب کاهش می یابد. اما برای توزیع پواسون به به دلیل پیچیدگی و وابسته بودن ریسک پسینی به متغیر تصادفی از شبیه سازی استفاده کردیم. در این مورد اگر پارامتر پیشینی α افزایش یابد با شرط ثابت بودن β ، حجم نمونه مطلوب کاهش می یابد و از طرف دیگر با شرط ثابت بودن α و افزایش β ، حجم نمونه مطلوب نیز افزایش می یابد.

به صورت $\delta = \frac{\beta+t}{\alpha+n}$ است. بنا بر این با استفاده از قضیه ۲.۲، تابع ریسک پسینی به صورت زیر به دست می آید:

$$PR = 1 - e \times \left(1 + \frac{1}{t+\beta}\right)^{-(t+\beta+1)}$$

حال با اضافه کردن ریسک پسینی به تابع هزینه (۱)، هزینه کل به صورت زیر به دست می آید:

$$TC(n) = c + cn + 1 - e \times \left(1 + \frac{1}{t+\beta}\right)^{-(t+\beta+1)}$$

که به T بستگی دارد. در این جا برای به دست آوردن حجم نمونه مطلوب باید $E(TC(n))$ یا به عبارت دیگر APR را به دست آورد که به دلیل شکل پیچیده ریسک پسینی، حل آن به روش آنالیزی امکان پذیر نیست؛ بنا بر این از شبیه سازی استفاده می کنیم. در جدول ۵ حجم نمونه مطلوب در برآورد پارامتر توزیع پواسون برای مقادیر مختلف پارامترهای پیشینی α و β آمده است. در این جدول مشاهده می شود که اگر پارامتر پیشینی α افزایش یابد با شرط ثابت بودن β ، حجم نمونه مطلوب کاهش می یابد و از طرف دیگر با شرط ثابت بودن α و افزایش β ، حجم نمونه مطلوب نیز افزایش می یابد. در شکل ۲ اندازه نمونه را در برابر هزینه نمونه گیری برای هر واحد، c با پارامترهای پیشینی $\alpha = 1$ و $\beta = 4$ رسم کرده ایم، که حجم نمونه مطلوب برابر با ۱۵ است.

مراجع

- [1] Adcock, C. J. (1988). A Bayesian approach to calculating sample size. *The Statistician*, **37**, 433-439.
- [2] Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd edn. New York: Springer.
- [3] Berger, J. O. (1980). *Statistical Decision Theory - Foundation Concepts and Methods*, Springer - Verlag.
- [4] Berger, J. O., Boukai, B. and Wang, Y. (1997). Unified frequentist and Bayesian testing of a precise hypothesis. *Statistical Science*, **12**, 133-160.
- [5] Goldstein, M. (1981). A Bayesian Criterion for Sample Size. *Annals of Statistics*, **9**, 670-672.
- [6] Grundy, P. M., Healy, M. J. R. and Rees, D. H. (1956). Economic choice of the amount of experimentation. *Journal of the Royal Statistical Society*, **A 18**, 32-48.
- [7] Hamilton, H. W. (1968). Confidence Properties of Bayesian Interval Estimates. *Journal of the Royal Statistical Society, Series B: Methodological*, **30**, 535-544.
- [8] Lindley, D. V. (1972). *Bayesian Statistics, a Review*. Philadelphia: Society for Industrial and Applied Mathematics.
- [9] Lindley, D. V. (1997a). The choice of sample size, *Statistician*, **2**, 129-138.

- [10] Maple 13 (1996-2009) Copyright ©Maplesoft, a division of Waterloo Maple Inc.
- [11] O'Hagan, A. and Stevens, J. W. (2001). Bayesian assessment of Sample size for clinical trails of cost-effectiveness. *Medical Decision Making*, **21**, 219-230.
- [12] Pezeshk, H. (2003). Bayesian techniques for Sample size determination in clinical trials: A short review. *Statistical Method in Medical Research*, **12(6)**, 489-504.
- [13] Pezeshk, H. and Gittins, J. C. (1999). Sample size Determination in Clinical Trials. *Student*, **3**, 19-26.
- [14] Pham-Gia, T. and Turkkan, N. (1992). Sample Size Determination in Bayesian Analysis (Disc: P399-404). *The Statistician: Journal of the Institute of Statisticians*, **41**, 389-397.
- [15] Raiffa, H. and Schlaifer, R. (1961). *Applied statistical Decision Theory*. Boston, Havard University Graduate School of Business Administration.
- [16] Sahu, S. K. and Smith, T. M. F. (2006). A Bayesian method of sample size determination with practical applications. *Journal of the Royal Statistical Society*, **A 169**, 235–253.
- [17] Saiful Islam, A. F. M. and Pettit, L. I. (2012). Bayesian sample size determination using linex loss and linear cost, *Communications in Statistics - Theory and Methods*, **41**, 223-240.
- [18] Spiring, F.A. and Yeung, A.S. (1998). A general class of loss functions with industrial applications, *Journal of Quality Technology*, **30**, 152-162.
- [19] Stallard, N. (1998). Sample size determination for phase II clinical trials based on Bayesian decision theory. *Biometrics*, **54**, 279-294.